REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
21-08-2012	Conference Proceeding	-

4. TITLE AND SUBTITLE
Algebraic Information Theory and Stochastic Resonance for Binary Input Binary Output Channels

5a. CONTRACT NUMBER
W911NF-11-1-0144

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
206022

6. AUTHORS
Ira S. Moskowitz, Paul Cotae, Pedro N. Safier

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
The University of the District of Columbia
Graduate Studies
University of the District of Columbia
Washington, DC 20008 -0000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
58962-NS-REP.4

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
In this paper, we analyze the information theoretic aspects of the phenomenon of stochastic resonance in terms of the algebraic properties of the channel matrix. The binary-input binary-output channel model under discussion is a threshold system. We offer an algebraic view of Kosko's Forbidden Interval

15. SUBJECT TERMS
Stochastic resonance, information theory, forbidden interval theorem.

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON	19b. TELEPHONE NUMBER
UU	UU	UU	UU		Paul Cotae	202-274-6290

Standard Form 298 (Rev 8/98)
Prescribed by ANSI Std. Z39.18

Report Title

Algebraic Information Theory and Stochastic Resonance for Binary Input Binary Output Channels

ABSTRACT

In this paper, we analyze the information theoretic aspects of the phenomenon of stochastic resonance in terms of the algebraic properties of the channel matrix. The binary-input binary-output channel model under discussion is a threshold system. We offer an algebraic view of Kosko's Forbidden Interval Theorem, and extend his result in certain cases.

Conference Name: Proceedings IEEE 46th Annual Conference on Information Sciences and Systems-CISS 2012, Princeton

Conference Date: March 21, 2012

Algebraic Information Theory and Stochastic Resonance for Binary-Input Binary-Output Channels

Ira S. Moskowitz
Center for High Assurance Computing
Code 5540, Naval Research Laboratory, Washington, DC 20375
Email: ira.moskowitz@nrl.navy.mil
&
Paul Cotae
Electrical Engineering Department
School of Engineering and Applied Sciences
University of the District of Columbia, Washington, DC 20008
Email: pcotae@udc.edu
&
Pedro N. Safier
S&J Solutions LLC
107 S. West St. PMB 509, Alexandria, VA 22314
Email: pedro-safier@sj-solutions.com

***Abstract*—In this paper, we analyze the information theoretic aspects of the phenomenon of stochastic resonance in terms of the algebraic properties of the channel matrix. The binary-input binary-output channel model under discussion is a threshold system. We offer an algebraic view of Kosko's Forbidden Interval Theorem, and extend his result in certain cases.**

***Index Terms*—Stochastic resonance, information theory, forbidden interval theorem.**

I. Introduction

In this paper, we analyze the information theoretic aspects of the phenomenon of stochastic resonance in terms of the algebraic properties of the channel matrix for a simple, but physically meaningful, threshold based system.

The communication channels under study in this paper are binary-input binary-output discrete memoryless channels [11]. That is, the channel is discrete, memoryless, and has a stationary channel matrix, the input random variable X has values $\{\iota_1, \iota_2\}$, and the output random variable Y has values $\{o_1, o_2\}$. The event $X = \iota_k$ is the event that the symbol ι_k is input to the channel, and the event $Y = o_k$ is the event that the channel output is the symbol o_k. We use the shorthand notation of a (2,2) channel, respectively.

We use the $(2, 2)$ channel to study the phenomena of stochastic resonance (SR) that arises in the discrete time threshold neuron model as first described by [4]. A nice introduction to this is given in [5, Sec. 2]. However, we closely follow the mathematics of [2] in this paper.

Algebraic information theory (AIT) is an approach to Shannon information theory based upon the algebraic properties, such as the determinant and eigenvalues, of the channel matrix (e.g. [9], [3]). This idea was briefly discussed in [13]. We view Kosko's FIT [5, Thm. 1], [6], [7] in light of AIT.

I-A. *Mutual Information and Capacity for (2,2) Channels*

The channel matrix represents the conditional probability relationships between the input and output symbols. That is, $a = P(o_1|\iota_1), \bar{a} = P(o_2|\iota_1), b = P(o_1|\iota_2)$, and $\bar{b} = P(o_2|\iota_2)$, where, in general, $\overline{(\cdot)} = 1-(\cdot)$. We have $P(X = \iota_1) = x_1$ and $P(X = \iota_2) = x_2, P(Y = o_1) = y_1$ and $P(Y = o_2) = y_2$. Since $x_2 = \overline{x_1}$ we simplify notation further and set $x := x_1$ and, similarly, $y := y_1$. We have the following equivalent expressions:

$$(y_1, y_2) = (x_1, x_2) \cdot \begin{pmatrix} a & \bar{a} \\ b & \bar{b} \end{pmatrix}, \ (y, \bar{y}) = (x, \bar{x}) \cdot \begin{pmatrix} a & \bar{a} \\ b & \bar{b} \end{pmatrix}. \quad (1)$$

and the channel matrix M, where

$$M := \begin{pmatrix} a & \bar{a} \\ b & \bar{b} \end{pmatrix}. \quad (2)$$

We see that the (2,2) channel is completely described by the channel matrix M. However, since M is a stochastic matrix, M itself is completely described by the 2-tuple (a, b). Thus, a (2,2) channel is uniquely identified with a point in the unit square $[0, 1] \times [0, 1]$.

Using standard Shannon information theory [15] we have the mutual information

$$I = H(Y) - H(Y|X) \quad (3)$$

where $H(Y)$ is the entropy[1] of the random variable Y, which for the (2,2) channel is the same as $h(y)$ and $H(Y|X)$ is the conditional entropy. Holding (a, b) fixed, we have that the

[1]Unless noted otherwise all logarithmic bases are base 2, where $\log(x)$ is the base 2 logarithm of x. If we wish to use the natural logarithm we will use the notation ln. The binary entropy function is denoted as $h(x) := -x\log(x)-(1-x)\log(x)$, and the natural binary entropy function is denoted as $h_e(x) := -x\ln(x) - (1-x)\ln(x)$.

mutual information given by Eq. (3) I is a function of x, setting $f(x) = (a-b)x + b$, we have that

$$I(x) = h\left(f(x)\right) - xh(a) - \bar{x}h(b). \tag{4}$$

The capacity C of the (2,2) channel determined by the parameters (a,b) is the well-defined continuous maximum of the mutual information ([16, Eq. 5], Ash [1, Eq. 3.3.5], or [8], [9])

$$\begin{aligned} \mathrm{C}(a,b) &= \frac{\bar{a}h(b)-\bar{b}h(a)}{a-b} + \log\left(1 + 2^{\frac{h(a)-h(b)}{a-b}}\right) \\ &= \log\left(2^{\frac{\bar{a}h(b)-\bar{b}h(a)}{a-b}} + 2^{\frac{bh(a)-ah(b)}{a-b}}\right) \end{aligned} \tag{5}$$

where $\mathrm{C}(a,a) := 0$. Note that capacity is zero if and only if $a = b$ (concavity arguments show that for all x, $I(x) \equiv 0$ iff $a = b$). One also has the symmetries

$$C(a,b) = C(b,a) = C(1-a,1-b) = C(1-b,1-a)\ . \tag{6}$$

In summary, $C(a,b)$ is defined and continuous for all $(a,b) \in [0,1]\times[0,1]$, is zero iff $a = b$, and is 1 if and only if $(a,b) = (1,0)$ or $(0,1)$, with the symmetries as illustrated in Figure 1.

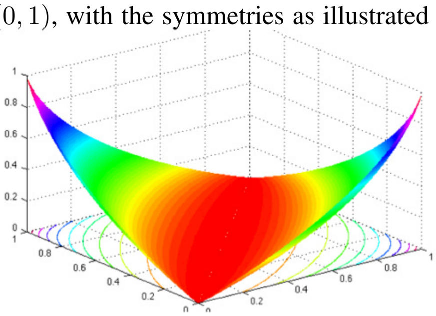


Fig. 1. Capacity as a function of (a, b).

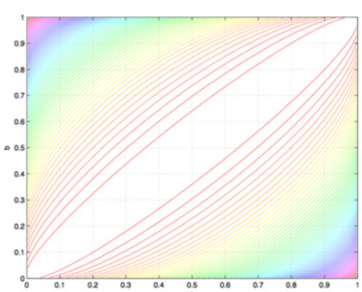


Fig. 2. Capacity level sets

The level sets of capacity can be seen in both Figures 1 & 2. To solve for b in terms of a, or visa versa, for a given level set does not seem to have a closed form solution due to the high non-linearity of the capacity expression involving logarithms and the binary entropy function. This makes working with them directly rather difficult, hence the approximations as developed in [11], [12].

II. THE $(2|2;\theta)$ THRESHOLD SYSTEM

The $(2|2;\theta)$ threshold system is a binary-input binary-output discrete time threshold system. We will not go into detail on the neuroscience background of (discrete time) threshold systems. We point the interested reader to [10]. We start, though, with an exposition as given by [2], then we normalize those results into the notation of [7]. We borrow freely from [13] in what follows. We use the notation $(\alpha,\beta|2;\theta)$ to specifically denote a $(2|2;\theta)$ threshold system with inputs α or β with an understood additive noise distribution. If we wish to make the noise distribution explicit we will write $(2|2;\theta;\mathfrak{N}_{\mathfrak{D}})$

The inputs to the threshold system are a signal $\mathfrak{X}$ taking on the values α or β. The (disturbance) noise is added to the transmission signal resulting in the received signal $\mathfrak{R}$:

$$\mathfrak{R} = \mathfrak{X} + \mathfrak{N}_{\mathfrak{D}}\ . \tag{7}$$

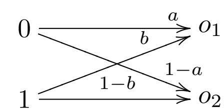
which is now thresholded resulting in Y

$$Y = \begin{cases} o_1, & \text{if } \mathfrak{R} \le \theta; \\ o_2 & \text{if } \mathfrak{R} > \theta. \end{cases} \tag{8}$$

II-A. *Shannon Model of the $(0,1|2;\theta;\mathcal{N}(\mu,\sigma^2))$ threshold system*

$\mathcal{N}(\mu,\sigma^2)$ is a normal distribution with mean μ and variance (power) σ^2.

$$M = \begin{pmatrix} a & 1-a \\ b & 1-b \end{pmatrix} \tag{9}$$



The inputs to the channel are 0 or 1, and the outputs are the two distinct symbols o_1 and o_2. The actual output values do not affect the information theoretic analysis of the problem (the input values do!).

Following[2] [2], we are letting $\mathfrak{N}_{\mathfrak{D}}$ be $\mathcal{N}(0,\sigma^2)$, with probability density function $f_\sigma(t) = \frac{1}{\sqrt{2\pi}\cdot\sigma}\exp(\frac{-t^2}{2\sigma^2})$, and cumulative distribution function $F_\sigma(x) = \int_{-\infty}^{x} f_\sigma(t)dt$. We denote the standard normal ($\mathcal{N}(0,1)$) cumulative distribution function as

$$\Phi(x) := \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{\frac{-t^2}{2}}dt\ . \tag{10}$$

Change of variables shows that $F_\sigma(x) = \Phi(x/\sigma)$. We take $\sigma > 0$, and look at the limiting situations as $\sigma \to 0+$.

The input random variable X represents the physical transmission of a signal $\mathfrak{X}$. In our situation $\mathfrak{X}$ takes on two distinct values, 0 or 1. The disturbance is added to the transmission signal resulting in the received signal $\mathfrak{R}$:

$$\mathfrak{R} = \mathfrak{X} + \mathfrak{N}_{\mathfrak{D}}\ . \tag{11}$$

The output random variable is determined as:

$$Y = \begin{cases} o_1, & \text{if } \mathfrak{R} \le \theta; \\ o_2 & \text{if } \mathfrak{R} > \theta. \end{cases} \tag{12}$$

Therefore, we easily have [2] that
$a = P(0+\mathcal{N}(0,\sigma^2) \le \theta) = P(\mathcal{N}(0,\sigma^2) \le \theta) = F_\sigma(\theta)$, and
$b = P(1+\mathcal{N}(0,\sigma^2) \le \theta) = P(\mathcal{N}(0,\sigma^2) \le \theta-1) = F_\sigma(\theta-1)$.

[2] In general, the noise need not have such a simple form, and is a subject of considerable research such as a general dynamical model in terms of stochastic differential equations ([14]). However, this paper is only concerned with the zero mean normal distribution ($\mathcal{N}(0,\sigma^2)$).

So, the channel matrix is of the form[3]

$$M = \begin{pmatrix} a & 1-a \\ b & 1-b \end{pmatrix} = \begin{pmatrix} F_\sigma(\theta) & \overline{F_\sigma(\theta)} \\ F_\sigma(\theta-1) & \overline{F_\sigma(\theta-1)} \end{pmatrix}$$
$$= \begin{pmatrix} \Phi\left(\frac{\theta}{\sigma}\right) & \overline{\Phi\left(\frac{\theta}{\sigma}\right)} \\ \Phi\left(\frac{\theta-1}{\sigma}\right) & \overline{\Phi\left(\frac{\theta-1}{\sigma}\right)} \end{pmatrix} . \tag{13}$$

So we have that the capacity, in closed form, is

$$C\left(F_\sigma(\theta), F_\sigma(\theta-1)\right) = \tag{14}$$
$$\log\left(2^{\frac{\overline{F_\sigma(\theta)}\cdot h(F_\sigma(\theta-1)) - \overline{F_\sigma(\theta-1)}\cdot h(F_\sigma(\theta))}{F_\sigma(\theta)-F_\sigma(\theta-1)}} + 2^{\frac{F_\sigma(\theta-1)\cdot h(F_\sigma(\theta)) - F_\sigma(\theta)\cdot h(F_\sigma(\theta-1))}{F_\sigma(\theta)-F_\sigma(\theta-1}}\right) .$$

Unfortunately, this does not seem to be a very tractable expression. This is where we turn to algebraic information theory. Keep in mind that

$$C\left(F_\sigma(\theta), F_\sigma(\theta-1)\right) = C\left(\Phi\left(\frac{\theta}{\sigma}\right), \Phi\left(\frac{\theta-1}{\sigma}\right)\right) . \tag{15}$$

III. Algebraic Information Theory

First, we review some notation and results from [9]. We say that a (2,2) channel is nonnegative if $a \geq b$, and it is positive if $a > b$. The negative channels are those with $a < b$, and the zero channels are those with $a = b$. The set of zero channels is identical to the set of channels with zero capacity. In [9], the directed complete partially ordered set (dcpo) of compact subintervals of the unit interval with the Scott topology $\mathbf{I}[0,1]$ is studied. The nonnegative (2,2) channel (a,b) with matrix $M = \begin{pmatrix} a & \bar{a} \\ b & \bar{b} \end{pmatrix}$ is uniquely identified (under a monoid isomorphism) with the interval $[b,a] \in \mathbf{I}[0,1]$.

Algebra fact 1

What is important for us is that [9, Thm. 4.9] tells us for nonnegative channels that

Theorem 3.1: (AIT ordering theorem)

$$a < a' \implies C(a,b) < C(a',b), \text{ and}$$
$$b' < b \implies C(a,b) < C(a,b') . \tag{16}$$

This seemingly obvious[4] result simply follows by taking the partial derivatives of the mutual information.

Algebra fact 2

Keep in mind that the determinant of the channel matrix is

$$\det M = a - b . \tag{17}$$

The following fundamental relation reveals much about capacity in terms of the determinant. It is shown in [11] that

$$\frac{(a-b)^2}{2\ln(2)} \leq C(a,b) \leq |a-b| . \tag{18}$$

[3]Recall $\bar{x} = 1 - x$.

[4]The authors of [9] were surprised that this result, to the best of their knowledge, did not exist in the literature prior to [9].

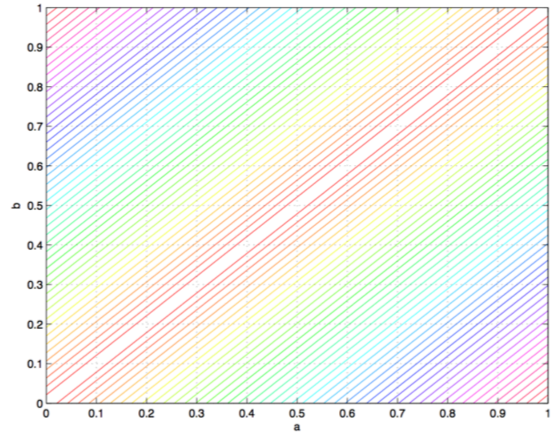


Fig. 3. $|a-b|$ level sets, compare to level sets of capacity in Fig. 2.

It is worth noting that the level sets of $\frac{(a-b)^2}{2\ln(2)}$ are identical, up to the level set , of the level sets of $|a-b|$ (Fig. 3). In essence what we are suggesting is a linearization of channel capacity. We wish to use this approach to discuss SR, and compare the linear approach to Kosko's FIT [5, Thm. 1], [6], [7].

Also keep in mind that for nonnegative channels

$$\frac{(a-b)^2}{2\ln(2)} \leq C(a,b) \leq a - b . \tag{19}$$

Therefore, for a nonnegative channel analyzing how $\det M = a - b$ changes, gives us guidance on the capacity behavior. Therefore, we can approximate the non-linear behavior of capacity by the linear behavior of $\det M$, in conjunction with the AIT ordering theorem. In this paper, we will concentrate on applying the AIT ordering theorem.

IV. Chapeau-Blondeau Results in Light of Algebraic Information Theory

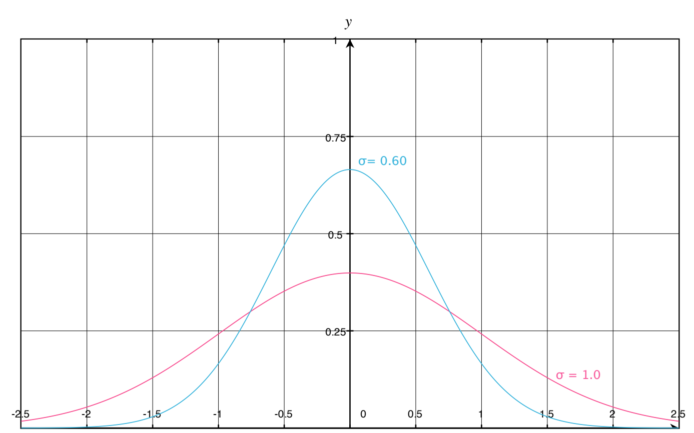


Fig. 4. Normal distributions $\mu = 0$, with $\sigma = 0.60$ (blue), $\sigma = 1.0$ (red).

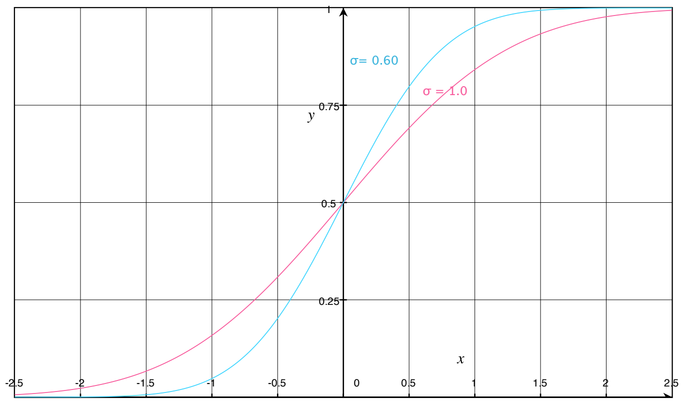


Fig. 5. $F_\sigma(\theta) = \Phi(\theta/\sigma)$ with $\sigma = 0.60$ (blue), $\sigma = 1.0$ (red).

It is interesting that there is ambiguity surrounding the formal definition of SR [10]. The essence of SR is that increasing noise can help the signal to noise ratio (SNR). Since variance is the power of a normal distribution signal, we adopt the following definition of SR.

For a $(0, 1|\mathcal{Z}; \theta; \mathcal{N}(0, \sigma^2))$ threshold system, we denote the Shannon channel capacity with θ fixed and variable σ as $C_\theta(\sigma)$.

Definition 4.1: We say that a $(0, 1|\mathcal{Z}; \theta; \mathcal{N}(0, \sigma^2))$ threshold system exhibits SR (as a function of σ) iff there exists $0 < \sigma_1$ and σ_2 such that $\sigma_1 < \sigma_2$ and $C_\theta(\sigma_1) < C_\theta(\sigma_2)$.

This goes along with the concept of adding noise (σ^2 is the power, and $\sigma = 0$ corresponds to no noise) to amplify the intended signal.

In [2], and in more detail in [13], it is graphically demonstrated and discussed that for $0 \leq \theta \leq 1$ there is no SR, but for $\theta < 0$ or $\theta > 1$ there is SR as σ goes from zero to infinity. Kosko's FIT [5], [6] can be used to prove this, but let us take an algebraic approach which we will later use for an algebraic proof of the FIT in general.

In Figure 6 (see [13, Fig. 3] we see that there is no SR for $\theta \in \{.8, .9, .95, .99, 1\}$, and there is SR with a single hump for $\theta \in \{1.01, 1.05, 1.1, 1.2\}$.

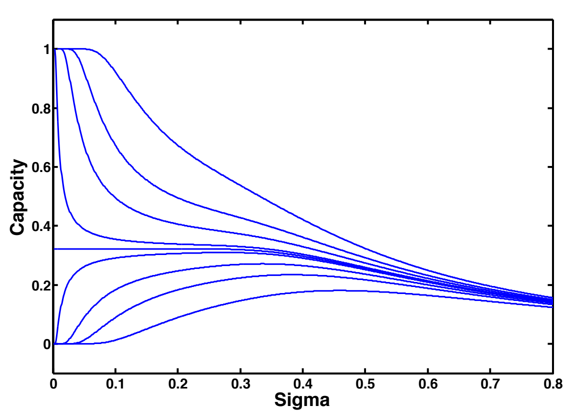


Fig. 6. Capacity, for $\theta \in \{.8, .9, .95, .99, 1, 1.01, 1.05, 1.1, 1.2\}$, as a function of σ.

Theorem 4.2: A $(0, 1|\mathcal{Z}; \theta; \mathcal{N}(0, \sigma^2))$ threshold system exhibits SR if and only if $\theta \notin [0, 1]$.

Proof: It is shown in [13, Thm1.1] that $C_{.5+d}(\sigma) = C_{.5-d}(\sigma)$. Therefore it suffices to show the threshold system does not exhibit SR for $\theta \in [.5, 1]$, and that it does exhibit SR for $\theta > 1$. (See Figs. 4 & 5).

ONLY IF part $\Leftarrow$

— For $\theta \in [.5, 1)$

Since $a = \Phi(\theta/\sigma)$ we see that since $\theta > 0$, a decreases as σ increases. On the other hand, $b = \Phi(\theta - 1)$ increases, since $\theta - 1 < 0$. Since a decreases and b increases, the AIT theorem tells us capacity decreases as σ increases.

— For $\theta = 1$

As above, a decreases as σ increases. However, $b = F_\sigma(0) \equiv 1/2$ for all σ. Since a decreases and b is constant the AIT theorem still tells us capacity decreases as σ increases.

Thus, there is no SR for $\theta \in [.5 : 1]$.

IF part $\Rightarrow$

— For $\theta > 1$

In this situation both a and b decrease as σ increases, so we cannot apply the AIT theorem. However, since $\Phi(\infty) = 1$, we see that as $\sigma \to 0^+$ that both $a \to 1$ and $b \to 1$, so capacity approaches zero. Since there exist $\sigma > 0$ such $a = F_\sigma(\theta) \neq F_\sigma(\theta - 1) = b$, we that that capacity (which is continuous) is increasing for some interval of σ.

Thus, there is SR for $\theta > 1$ ■

Corollary 4.1: For $\theta \in [.5, 1)$, capacity decreases from a limiting value of 1 down to a limiting value of 0, and for $\theta = 1$, capacity decreases from the limiting Z-channel capacity of $\approx .3219$ down to a limiting value of 0. In addition, for $\theta > 1$ the plot of capacity will exhibit at least one hump (see Figure 6).

Proof:

— For $\theta \in [.5, 1)$

As $\sigma \to 0^+, (a, b) \to (1, 0)$, so $C_\theta(0) \to 1$.

As $\sigma \to \infty, (a, b) \to (1/2, 1/2)$, so $C_\theta(\infty) \to 0$.

— For $\theta = 1$

As $\sigma \to 0^+, (a, b) = (1, 1/2)$, so $C_\theta(0) \to \approx .3219$.

As $\sigma \to \infty, (a, b) \to (1/2, 1/2)$, so $C_\theta(\infty) \to 0$.

— For $\theta > 1$

As above, as $\sigma \to 0^+$, both $a \to 1$ and $b \to 1$, so capacity approaches zero. As $\sigma \to \infty$ both $a \to 1/2$ and $b \to 1/2$, so capacity approaches zero. Therefore, there is at least one "hump" in the plot. ■

V. Kosko Bipolar Threshold System

Say we have a $(0, 1|\mathcal{Z}; \theta'; \mathcal{N}(0, \sigma'^2))$ threshold system. Consider a normalization such that $\theta = 2\theta' - 1$ and $\sigma = 2\sigma'$, and let the channel inputs {-1,1} replace {0,1} respectively.

The behavior of $(0, 1|\mathcal{Z}; \theta'; \mathcal{N}(0, \sigma'^2))$ threshold system is the same as a $(-1, 1|\mathcal{Z}; \theta; \mathcal{N}(0, \sigma^2))$ threshold system, where $\theta = 2\theta' - 1$ and $\sigma = 2\sigma'$. This means we identify the closed interval $[0, 1]$ with $[-1, 1]$ under the mapping $x \to 2x - 1$ and we analyze $(-1,1|2;2\theta'-1;\mathcal{N}(0,2\sigma'^2))$ and the SR regions are the same.

So for a $(-1,1|2;2\theta'-1;\mathcal{N}(0,2\sigma'^2))$ threshold system we have
$a = P(-1+\mathcal{N}(0,\sigma^2) \leq \theta) = P(\mathcal{N}(0,\sigma^2) \leq \theta+1) = F_\sigma(\theta+1) = F_{2\sigma'}(2\theta') = \Phi(\theta'/\sigma')$, and
$b = P(1+\mathcal{N}(0,\sigma^2) \leq \theta) = P(\mathcal{N}(0,\sigma^2) \leq \theta-1) = F_\sigma(\theta-1) = F_{2\sigma'}(2\theta'-2) = \Phi((\theta'-1)/\sigma')$.
So, the channel matrix is of the form

$$M = \begin{pmatrix} a & 1-a \\ b & 1-b \end{pmatrix} = \begin{pmatrix} F_\sigma(\theta+1) & \overline{F_\sigma(\theta+1)} \\ F_\sigma(\theta-1) & \overline{F_\sigma(\theta-1)} \end{pmatrix} = \begin{pmatrix} \Phi\left(\frac{\theta'}{\sigma'}\right) & \overline{\Phi\left(\frac{\theta'}{\sigma'}\right)} \\ \Phi\left(\frac{\theta'-1}{\sigma'}\right) & \overline{\Phi\left(\frac{\theta'-1}{\sigma'}\right)} \end{pmatrix}. \tag{20}$$

Therefore, we see that the channel matrix for the $(-1,1|2;\theta;\mathcal{N}(0,\sigma^2))$ threshold system is the same as the channel matrix for the $(0,1|2;\theta';\mathcal{N}(0,\sigma'^2))$, up to a normalization. In fact, there is nothing special about ± 1, so let us use $\pm A, A>0$ for the X inputs to the system (and channel).

For a $(-A,A|2;\theta;\mathcal{N}(0,\sigma^2))$ threshold system, we denote the Shannon channel capacity with θ fixed and variable σ as $C_\theta(\sigma)$.

Definition 5.1: We say that a $(-A,A|2;\theta;\mathcal{N}(0,\sigma^2))$ threshold system exhibits SR (as a function of σ) iff there exists $0<\sigma_1$ and σ_2 such that $\sigma_1<\sigma_2$ and $C_\theta(\sigma_1) < C_\theta(\sigma_2)$.

Theorem 5.2: If we map the interval $[0,1] \to [-A,A]$ under the affine map $x \to A(2x-1)$ and set $\theta := A(2\theta'-1)$, and $\sigma := 2A\sigma'$, then the channel matrices, as functions of σ' and θ', for the $(0,1|2;\theta';\mathcal{N}(0,\sigma'^2))$ and $(-A,A|2;\theta;\mathcal{N}(0,\sigma^2))$ threshold systems are the same. Furthermore, the $(-A,A|2;\theta;\mathcal{N}(0,\sigma^2))$ threshold system exhibits SR if and only if $\theta \notin [-A,A]$, and its capacity, as a function of σ, exhibits a symmetry about 0.

Proof: For the $(-A,A|2;\theta;\mathcal{N}(0,\sigma^2))$ threshold system
$a = P(-A+\mathcal{N}(0,\sigma^2) \leq \theta) = P(\mathcal{N}(0,\sigma^2) \leq \theta+A) = F_\sigma(\theta+A) = F_{2A\sigma'}(2A\theta') = \Phi(\theta'/\sigma')$, and
$b = P(A+\mathcal{N}(0,\sigma^2) \leq \theta) = P(\mathcal{N}(0,\sigma^2) \leq \theta-A)$
$= F_\sigma(\theta-A) = F_{2A\sigma'}(2A\theta'-2A) = \Phi((\theta'-1)/\sigma')$.
So, by Thm 4.2, the $(-A,A|2;\theta;\mathcal{N}(0,\sigma^2))$ threshold system exhibits SR if and only if $\theta' \notin [0,1]$, which, since $\theta' = \frac{\theta+A}{2A}$, is equivalent to $\theta \notin [-A,A]$.[5] ■

V-A. Non-zero Mean

Up to now we have been looking at how θ relates to the interval $[-A,A]$ with additive disturbance $\mathcal{N}(0,\sigma^2))$. Now let us consider additive disturbance with non-zero mean, that is $\mathcal{N}(\mu,\sigma^2))$. For a $(-A,A|2;\theta;\mathcal{N}(\mu,\sigma^2))$ threshold system we denote the Shannon channel capacity with θ fixed and variable σ as $C_\theta(\sigma)$.

[5]At this point we are not concerned with which standard deviation value maximizes capacity, so we do not concern ourselves with the distinction between σ and σ'. We include this footnote to assure the reader that we have not forgotten about the normalization.

Definition 5.3: We say that a $(-A,A|2;\theta;\mathcal{N}(\mu,\sigma^2))$ threshold system exhibits SR (as a function of σ) iff there exists $0<\sigma_1$ and σ_2 such that $\sigma_1<\sigma_2$ and $C_\theta(\sigma_1) < C_\theta(\sigma_2)$.

Consider the map $x \to A(2x-1)+\mu$ and set $\theta := A(2\theta'-1)+\mu$, and $\sigma := 2A\sigma'$.

So, for a $(-A,A|2;\theta;\mathcal{N}(\mu,\sigma^2))$ threshold system, we have that $a = P(-A+\mathcal{N}(\mu,\sigma^2) \leq \theta) = P(\mathcal{N}(\mu,\sigma^2) \leq \theta+A) = \Phi\left(\frac{\theta+A-\mu}{\sigma}\right) = \Phi(\theta'/\sigma')$, and similarly $b = \Phi\left(\frac{\theta-A-\mu}{\sigma}\right) = \Phi((\theta'-1)/\sigma')$. So there will be SR if and only if $\theta' \notin [0,1]$. However, since $\theta' = \frac{\theta+A-\mu}{2A}$, this is equivalent to

Theorem 5.4: The $(-A,A|2;\theta;\mathcal{N}(\mu,\sigma^2))$ threshold system exhibits SR if and only if $\theta \notin [\mu-A,\mu+A]$, or equivalently, $\mu \notin [\theta-A,\theta+A]$.

We consider Kosko's FIT [5, Thm. 1], [6], [7] in light of AIT. Note, his region of interest is $\mu \notin [\theta-A,\theta+A]$.

V-B. Kosko Formulation

We use [6, Thm. 1.1, Thm 1.2] for Kosko's FIT.

1) Instead of our Eq. (8), Kosko has the slightly different
$$Y = \begin{cases} o_1, & \text{if } \Re < \theta; \\ o_2 & \text{if } \Re \geq \theta. \end{cases} \tag{21}$$
This has the effect of changing the closed interval $[\theta-A,\theta+A]$ to the open interval $(\theta-A,\theta+A)$.

2) Kosko requires the input signal to be subthreshold; that is $A<\theta$. We do not.

3) We restricted the disturbance noise to be Gaussian, Kosko allows much more freedom, and in fact studies non-finite variances. We will address this in future work. Please note that the "If" part of our proof is essentially the same as Kosko's so there should be no problem with non-Gaussian distributions (at least those with finite variance).

4) The "Only If" part our our proof relies on AIT. We show that the capacity is monotone decreasing. Kosko shows that capacity decreases from the maximal value (obtained from the limit when $\sigma \to 0+$), but he does not show that there are no local maxima. This is an important distinction. This is why our generalized definition of SR is as given in Def. 5.3.

VI. Capacity Approximations

In [11], [12], [13] we showed how capacity of the channels under study in this paper can be simply approximated and bounded by simple functions of the determinant of the channel matrix. For the $(-A,A|2;\theta;\mathcal{N}(\mu,\sigma^2))$ threshold system we have:

$$\frac{1}{2\ln(2)}\left(\Phi\left(\frac{\theta+A-\mu}{\sigma}\right) - \Phi\left(\frac{\theta-A-\mu}{\sigma}\right)\right)^2 \leq C_\sigma(\theta) \leq \Phi\left(\frac{\theta+A-\mu}{\sigma}\right) - \Phi\left(\frac{\theta-A-\mu}{\sigma}\right). \tag{22}$$

We want to see when these lower and upper approximations exhibit SR. That is simply done by taking the first derivative with respect to σ and setting it to zero. Since $\frac{d}{d\sigma}k(a-b)^2 = 2k(a-b) \cdot \frac{d}{d\sigma}(a-b)$, and $a > b$, we see that it suffices to consider $\frac{d}{d\sigma}\left(\Phi\left(\frac{\theta+A-\mu}{\sigma}\right) - \Phi\left(\frac{\theta-A-\mu}{\sigma}\right)\right) = 0$.

$$\frac{d}{d\sigma}\left(\Phi\left(\frac{\theta+A-\mu}{\sigma}\right) - \Phi\left(\frac{\theta-A-\mu}{\sigma}\right)\right) =$$
$$\frac{d}{d\sigma}\left(\frac{1}{\sqrt{2\pi}}\int_{\frac{\theta-A-\mu}{\sigma}}^{\frac{\theta+A-\mu}{\sigma}} e^{-\frac{t^2}{2}}\,dt\right) =$$
$$\frac{1}{\sqrt{2\pi}}\left[\frac{-(\theta+A-\mu)}{\sigma^2}e^{\frac{-\left(\frac{\theta+A-\mu}{\sigma}\right)^2}{2}} - \frac{-(\theta-A-\mu)}{\sigma^2}e^{\frac{-\left(\frac{\theta-A-\mu}{\sigma}\right)^2}{2}}\right].$$

Setting the above to zero we arrive at:

$$\frac{\theta+A-\mu}{\theta-A-\mu} = e^{\frac{(\theta+A-\mu)^2-(\theta-A-\mu)^2}{2\sigma^2}} \tag{23}$$
$$\frac{\theta-(\mu-A)}{\theta-(\mu+A)} = e^{\frac{4A(\theta-\mu)}{2\sigma^2}} \tag{24}$$

Provided that the LHS of Eq. (24) is positive, which is true iff $\theta \notin [\mu-A, \mu+A]$, we arrive at

$$\ln\left(\frac{\theta+A-\mu}{\theta-A-\mu}\right) = \frac{4A(\theta-\mu)}{2\sigma^2} \tag{25}$$
$$\sigma^2 = \frac{2A(\theta-\mu)}{\ln\left(\frac{\theta-(\mu-A)}{\theta-(\mu+A)}\right)}. \tag{26}$$

Provided that the RHS if the above is positive, we have

$$\sigma = \sqrt{\frac{2A(\theta-\mu)}{\ln\left(\frac{\theta-(\mu-A)}{\theta-(\mu+A)}\right)}}. \tag{27}$$

If $\theta > \mu + A)$ then $\frac{\theta-(\mu-A)}{\theta-(\mu+A)} > 1$, if $\theta < \mu - A)$ then $0 < \frac{\theta-(\mu-A)}{\theta-(\mu+A)} < 1$. Therefore Eq. (27) has a solution iff $\theta \notin [\mu-A, \mu+A]$.

Definition 6.1: We call the σ derived from Eq. (27) the critical sigma approximation, and denote it as $\sigma_{\widetilde{sr}}$.

We note that in [13] we have given graphical evidence for how well the algebraic determinant bounds approximate capacity, and also how well $\sigma_{\widetilde{sr}}$ corresponds with the actual sigma that locally maximizes capacity.

What is most interesting is that one can obtain $\sigma_{\widetilde{sr}}$ iff the constraints of the FIT hold, that is $\theta \notin [\mu-A, \mu+A]$, or equivalently $\mu \notin [\theta-A, \theta+A]$. Therefore, our approximating capacity by the determinant bounds gives us another form of the FIT, which is derived solely from the algebraic properties of the channel matrix. This holds promise for approximating when more complex threshold systems exhibit SR.

VII. Conclusion and Future Work

We have amplified the definition of stochastic resonance, and shown how an algebraic approach to capacity and the associated issues of SR is a valid approach. We have given an alebraic proof for part of the FIT. We have shown how determinant bounds can easily give us the regions where SR occur.

In future work, we will analyze non-Gaussian distributions, and see if our algebraic approach can be applied to more complex systems. In particular, we want to see if our algebraic results for (2,3) channels [12] can be applied to similar threshold systems.

Acknowledgments

The authors would like to thank to Ruth Irene and Perkei Avos for their insightful comments. The second author wishes to acknowledge the support of the Department of Defense Grant W911NF-11-1-0144. The authors are also grateful to Bart Kosko for his helpful email prior to the writing of this paper.

References

[1] Robert B. Ash. *Information Theory*. Dover, 1990.

[2] François Chapeau-Blondeau. Noise-enhanced capacity via stochastic resonance in an asymmetric binary channel. *Physical Review E*, 55(2):2016 – 2019, 1997.

[3] Paul Cotae, Ira S. Moskowitz, and Myong H. Kang. Eigenvalue characterization of the capacity of discrete memoryless channels with invertible channel matrices. In *Proc. 44th Annual Conf. on Information Sciences and Systems, CISS 2010*, Princeton, NJ, USA, March 2010.

[4] J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the USA*, 79:2554–2558, April 1982.

[5] Bart Kosko and Sanya Mitiam. Stochastic resonance in noisy threshold neurons. *Neural Networks*, 16:755–761, 2003.

[6] Bart Kosko and Sanya Mitiam. Robust stochastic resonance for simple threshold neurons. *Physical Review E*, 70(3):0331911–1—0331911–10, 2004.

[7] Bart Kosko, Sanya Mitiam, Ashok Patel, and Mark M. Wilde. Applications of forbidden interval theorem in stochastic resonance. In *Applications of Nonlinear Dynamics, Understanding Complex Systems*, pages 71–89. Springer-Verlag, 2009.

[8] Keye Martin and Ira S. Moskowitz. Noisy timing channels with binary inputs and outputs. In Camenisch et al., editor, *Information Hiding, 8th International Workshop*, pages 124–144, Alexandria, VA, USA, July 2006. Springer-Verlag, LNCS 4437.

[9] Keye Martin, Ira S. Moskowitz, and Gerard Allwein. Algebraic information theory for binary channels. *Theoretical Computer Science*, 411:1918–1927, April 2010. Revised version: Proc. MFPS XXII, 2006.

[10] Mark D. McDonnell, Nigel G. Stocks, Charles E.M. Pearce, and Derek Abbott. *Stochastic Resonance: From Suprathreshold Stochastic Resonance to Stochastic Signal Quantization*. Cambridge, 2008.

[11] Ira S. Moskowitz. An approximation of the capacity of a simple channel. In *Proc. 43rd Annual Conf. on Information Sciences and Systems, CISS 2009*, pages 164–169, Baltimore, MD, USA, March 2009.

[12] Ira S. Moskowitz. Approximations for the capacity of binary input discrete memoryless channels. In *Proc. 44th Annual Conf. on Information Sciences and Systems, CISS 2010*, Princeton, NJ, USA, March 2010.

[13] Ira S. Moskowitz, Paul Cotae, Pedro N. Safier, and Daniel L. Kang. Capacity bounds and stochastic resonance for binary input binary output channels. In *to appear — Proc, Computing, Communications & Applications Conference, ComComAP 2012*. IEEE, 2012.

[14] Ashok Patel and Bart Kosko. Stochastic resonance in continuous and spiking neuron models with Levy noise. *IEEE Transactions on Neural Networks*, 19(12):71–89, 2009.

[15] Claude E. Shannon. A mathematical theory of communication. *Bell Systems Technical Journal*, 27:379–423,623–656, 1948.

[16] Richard A. Silverman. On binary channels and their cascades. *IRE Transactions on Information Theory*, 1(3):19–27, May 1955.